AD-A063 983 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
THE RELATION BETWEEN STATISTICAL DECISION THEORY AND APPROXIMAT--ETC(U)
OCT 78 B HARRIS, G HEINDL DAAG29-75-C-0024
UNCLASSIFIED MSR-TSR-1890 NL

1 OF 1
ADA
063983

END
DATE
FILMED
4-79
DDC

10 NW

9

MRC Technical Summary Report #1890

6

THE RELATION BETWEEN STATISTICAL DECISION THEORY AND APPROXIMATION THEORY

10

Bernard Harris ~~and~~ Gerhard Heindl

LEVEL II

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

11 Oct[redacted]978

12 14p.

Received September 28, 1978

15 DAAG29-75-C-ΦΦ24

14 MSR-TSR-189Φ

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709

JOB

78 12 27 068

221 200

THE RELATION BETWEEN STATISTICAL DECISION THEORY

AND APPROXIMATION THEORY

Bernard Harris* and Gerhard Heindl**

Technical Summary Report #1890
October 1978

ABSTRACT

This paper extends previous results by the first author (see for example, "Mathematical Models for Statistical Decision Theory" in Optimizing Methods in Statistics, J. S. Rustagi, Ed., Academic Press, 1971).

The approximation theory model describes a class of optimality principles in statistical decision theory as follows. Let S be the risk set of a statistical decision problem, that is, $S = \{R_\varphi(\theta), \theta \in \Theta, \varphi \in \Phi\}$ where Φ is the collection of randomized decision procedures, Θ is the parameter space and $R_\varphi(\theta)$ is the risk function of the statistical decision procedure φ. We interpret S as a set in the normed linear space L. Let $v = v(\theta)$ satisfy $v(\theta) \leq R_\varphi(\theta)$ for all $\varphi \in \Phi$ and all $\theta \in \Theta$. Then $s_0 \in S$ is said to be (v,L) optimal if $\|s_0 - v\| \leq \|s - v\|$ for all $s \in S$.

It is easily seen that many well-known optimality principles of statistics are of this type, such as Bayes rules and minimax rules.

In this paper, characterization theorems for this class of optimality principles are given.

AMS(MOS) Subject Classification - 62C05

Key Words: Decision theory
Approximation theory

Work Unit No. 4 - Probability, Statistics and Combinatorics

---

*Mathematics Research Center, University of Wisconsin, Madison, Wisconsin 53706.

**Mathematisches Institut der Technischen Universität München.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

78 12 27 068

SIGNIFICANCE AND EXPLANATION

This paper provides basic results for the problem of decision making under uncertainty. Such problems arise in virtually every area of human endeavor, such as the social sciences, biological sciences and physical sciences.

This paper characterizes optimality principles in decision theory by reformulating them as problems in approximation theory. Connections can be made between "best" approximations and "best" decision procedures. It is shown that many results in approximation theory have interpretations as results in decision theory, and conversely many results in statistical decision theory can be reformulated as results in approximation theory.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

THE RELATION BETWEEN STATISTICAL DECISION AND APPROXIMATION THEORY

Bernard Harris[*] and Gerhard Heindl[**]

1. Introduction and Summary. The purpose of this paper is to further explore the relationship between statistical decision theory and approximation theory initially presented in B. Harris [5]. As indicated there, many results of approximation theory, when suitably modified, have interpretations as results in statistical decision theory. Conversely, many results of statistical decision theory are also capable of reformulation as results in approximation theory.

In this paper, some results of approximation theory will be modified, reformulated and reinterpreted as results of statistical decision theory.

Let $\Theta \neq \phi$ be a given set with elements θ. Further let Φ be a non-empty family of probability measures defined on B_D, a σ-algebra of subsets of a set D. Thus, we are given the family of measure spaces (D, B_D, φ), $\varphi \in \Phi$. For each $\varphi \in \Phi$, let R_φ be a mapping from Θ into the extended real numbers. Let $R = \{R_\varphi, \varphi \in \Phi\}$, that is, R is a family of extended real valued functions of θ, $\theta \in \Theta$. It will be assumed that $R_\varphi(\theta)$ is *uniformly bounded* from below. In other words, there exists a real number M such that

$$(1) \qquad R_\varphi(\theta) \geq M, \quad \text{for all } \varphi \in \Phi \text{ and all } \theta \in \Theta .$$

In addition we also require that R be a convex set, that is, for every pair of functions $R_1, R_2 \in R$ and every λ, $0 \leq \lambda \leq 1$,

$$R_\lambda(1,2) = \lambda R_1 + (1-\lambda) R_2 \in R .$$

In the usual terminology employed in statistics, Θ is the parameter space and θ is called a parameter. R is the risk set of a statistical decision problem and R_φ is called a risk function. The reader is referred to standard treatises on statistical decision theory, such as D. Blackwell and M. A. Girshick [1] or T. S. Ferguson [4] for a detailed description of the manner in which the risk set is determined by a statistical decision problem.

[*] Mathematics Research Center, University of Wisconsin, Madison, Wisconsin 53706.

[**] Mathematisches Institut der Technischen Universität München.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

Within the framework described above a statistical decision problem may be described as follows. The statistician selects a $\varphi \in \Phi$ and the performance of the φ that he selects is described by the risk function R_φ. Consequently optimality in statistics may be roughly described by the intuitive statement; "φ is optimal if $R_\varphi(\theta)$ is as small as possible". However, in general, there is no $\varphi_0 \in \Phi$ such that $R_{\varphi_0}(\theta) \leq R_\varphi(\theta)$ for all θ and all $\varphi \in \Phi$. Consequently, there is no single optimality principle which enjoys universal acceptance. In a previous article on optimality principles in statistics, Harris [5] proposed a class of optimality principles called $(v,\mathcal{L})$ optimality, which we will now define.

Let $\mathcal{L}$ be a normed linear space with a partial ordering. Assume that for $0 \leq x \leq y$, $x,y \in \mathcal{L}$, we have $\|x\| \leq \|y\|$. We refer to this property by saying that the space $\mathcal{L}$ is endowed with a monotonic norm. Let $S_\mathcal{L} = \mathcal{R} \cap \mathcal{L}$, that is, the restriction of the risk set to those elements with finite norms. To avoid trivial situations, we will assume that $S_\mathcal{L} \neq \phi$. Let v be a distinguished point in $\mathcal{L}$ satisfying $v \leq R$, i.e., $v \equiv v(\theta) \leq R_\varphi(\theta)$ for all $\varphi \in \Phi$. The partial ordering in $\mathcal{L}$ will be defined as the partial ordering on the risk functions $R_\varphi(\theta)$ induced by admissibility. Thus we identify the points s in $S_\mathcal{L}$ with the risk functions, so that $s_1 \leq s_2$ means $R_{\varphi_1}(\theta) \leq R_{\varphi_2}(\theta)$ for all θ, where s_1 is identified with R_{φ_1} and s_2 is identified with R_{φ_2}. Similarly, $s_1 < s_2$ means $R_{\varphi_1}(\theta) \leq R_{\varphi_2}(\theta)$ for all θ and there exists a $\theta_0 \in \Theta$ such that $R_{\varphi_1}(\theta_0) < R_{\varphi_2}(\theta_0)$. Thus, $v \leq R$ insures $v \leq S_\mathcal{L}$.

Then, with this identification, we say that s_0 (equivalently φ_0) is $(v,\mathcal{L})$ optimal if $\|s_0 - v\| \leq \|s - v\|$ for all $s \in S_\mathcal{L}$.

It is easily seen that appropriate specifications of v and $\mathcal{L}$ give rise to various standard optimality principles of statistical theory, such as Bayes rules, minimax rules and minimax regret rules. Specifically, let $v = v(\theta) \equiv M$ and let τ be a specified prior probability measure. Then, if we take $\mathcal{L} = L_1(\mathcal{X}, \mathcal{B}_\mathcal{X}, \tau)$, that is the space of τ-integrable functions, then the assertion that s_0 is $(v,\mathcal{L})$ optimal is equivalent

to the assertion that s_0 is Bayes against τ. One can provide the following intuitive justification for the notion of (v,L) optimality. Since the partial ordering on L has been identified with admissibility, one can say that s_1 is "at least as good" as s_2 whenever $s_1 \leq s_2$ and s_1 "is better than s_2" whenever $s_1 < s_2$. The set S_L represents the totality of decision procedures to be considered by the statistician. Thus $v \leq R$ says that v is "at least as good" as any alternative available to the statistician. Hence v should be regarded as an ideal decision procedure. Naturally v need not be in S_L; v should be regarded as what you would do if you were able to, such as if you had "perfect information". Hence, it is natural to seek to come as close to v as possible and thus the idea of minimizing the norm from v to S_L. The distance from v to S_L may be regarded as a measurement of the "uncertainty".

In approximation theory a point x and a set S in a normed linear space L are specified. $s_0 \in S$ is said to be a best approximation to x from the set S if $\|s_0 - x\| \leq \|s - x\|$ for all $s \in S$. In many applications S is a linear subspace of L or a convex set in L. Extensive discussions of the theory of best approximations may be found in R. B. Holmes [7] or I. Singer [10]. However, if we add the additional conditions that L has a partial ordering and a monotonic norm and that $v \leq S$, where S is a convex set, then this becomes precisely the notion of (v,L) optimality in statistical decision theory. Thus (v,L) optimality may be interpreted as best one-sided approximation from a distinguished point to a convex set in a normed linear space with a monotonic norm.

We now show how methods of approximation theory may be used to obtain characterizations of (v,L) optimality.

2. Characterization of (v,L) optimality.

In this section we present some theorems which characterize (v,L) optimal decision procedures.

Theorem 1. Let $L \neq \{0\}$ be a partially ordered normed linear space of real valued functions of θ, $\theta \in \Theta$, equipped with a monotonic norm. Let S be a given convex set

in L and let $v \in L$ with $v \leq S$. Then $s_0 \in S$ is (v,L) optimal if and only if there is a linear functional $L \in L^*$, the topological dual space of L, such that (1) $\|L\| = 1$, (2) $L \in K^*$, the dual of the positive cone K in L, (3) $L(s_0) \leq L(s)$, $s \in S$, (4) $L(s_0 - v) = \|s_0 - v\|$.

<u>Proof</u>. <u>Sufficiency</u>. $\|s_0 - v\| = L(s_0 - v) \leq L(s - v) \leq \|L\| \, \|s - v\| = \|s - v\|$ for every $s \in S$ and hence s_0 is (v,L) optimal.

<u>Necessity</u>. K is a non-empty cone with 0 as vertex. We first show that $\bar{K} \neq L$. If $K = \{0\}$, then $K = \bar{K}$ and $\bar{K} \neq L$ since $L \neq \{0\}$. If $K \neq \{0\}$, then there exists an $x \geq 0$, $x \neq 0$. Now if $\bar{K} = L$, we must have $-x \in \bar{K}$ and there is a sequence $\{y_k\}$ with $y_k \geq 0$ for all k and $\lim_{k\to\infty} y_k = -x$. Then $0 \leq x \leq x + y_k$ and $0 \leq \|x\| \leq \|x + y_k\|$, $k = 1,2,\ldots$. Since $\lim_{k\to\infty} \|x + y_k\| = 0$, we have $\|x\| = 0$ and $x = 0$.

We now divide the balance of the proof into two parts, treating separately the cases $v \in \bar{S}$ and $v \notin \bar{S}$.

If $v \in \bar{S}$, then $\|v - s_0\| = 0$ and $v = s_0$. Then since $\bar{K} \neq L$, there is a linear functional $L \in L^*$, $L \neq 0$ such that $L(x) \geq 0$ for all $x \in \bar{K}$. Without loss of generality, we can set $\|L\| = 1$. Then for every $s \in S$, $s - s_0 = s - v \geq 0$ and $s - v \in K \subset \bar{K}$, which implies $L(s - v) \geq L(s_0 - v) = 0$; hence $L(s) \geq L(s_0)$.

Now suppose $v \notin \bar{S}$. Then $\|s_0 - v\| = d > 0$. Let $B(v,d) = \{x \in L : \|x - v\| < d\}$ and let $\hat{S} = \{s + x : s \in S, x \in K\}$. $\hat{S}$ is obviously convex since S and K are convex. Then for $z \in \hat{S}$, there exists $s \in S$ and $x \geq 0$ such that $0 \leq s - v \leq s - v + x = z - v$. Hence $\|s - v\| \leq \|z - v\|$ and $z \notin B(v,d)$. Consequently, $\hat{S} \cap B(v,d) = \phi$ and there exists a separating hyperplane for $\hat{S}$ and $B(v,d)$. That is, there exists an $L \in L^*$ and a real constant c such that $L \neq 0$ and

$$\begin{cases} L(x) \leq c, & x \in B(v,d) \\ L(z) \geq c, & z \in \hat{S}\ . \end{cases}$$

Further, we can set $\|L\| = 1$. Clearly $s_0 \in \overline{B(v,d)}$; also $s_0 \in \hat{S}$ and $S \subset \hat{S}$. Therefore $L(s) \geq c$ for all $s \in S$ and $L(s_0) = c$, establishing $L(s) \leq L(s_0)$. For $x \in K$, $s_0 + x \in \hat{S}$ and hence $L(s_0 + x) \geq L(s_0)$, which implies $L(x) \geq 0$.

Finally, for any z with $\|z\| \le 1$, $\|v + dz - v\| = d\|z\| \le d$ so that $v + dz \in \overline{B(v,d)}$. Consequently, $L(v + dz) \le L(s_0) = c$. Similarly $L(dz) \le L(s_0 - v)$ and thus $L(z) \le d^{-1}L(s_0 - v)$. Hence $\|L\| \le d^{-1}L(s_0 - v)$ and $d \le L(s_0 - v)$ $\le \|L\| \, \|s_0 - v\| = d$, establishing the theorem.

A special case of Theorem one is a well-known result in statistics and is given by the following corollary.

Corollary 1. Let Θ be a compact Hausdorff space and let $S = \{R_\varphi(\theta), \varphi \in \Phi\}$ be a convex set of continuous functions of θ, uniformly bounded from below. Then $s_0 \in S$ is minimax if and only if there exists a least favorable distribution τ_0 and s_0 is Bayes against τ_0.

Proof. Let $L = C_{[\Theta]}$, the space of continuous real valued functions on Θ with $\|f\| = \sup_{\theta\in\Theta} |f(\theta)|$ for $f \in L$. Then L^* is isometric to the set of regular countably additive set functions on the sets of the Borel σ-algebra of Θ (see N. L. Dunford and J. T. Schwartz [3], p. 265.). For $L \in L^*$, $\|L\|$ is the total variation of the set function L. Specializing Theorem 1 to this case, the conditions $L \ge 0$ and $\|L\| = 1$ establish that L is representable by a probability measure τ_0 on Θ. $L(s_0) \le L(s)$ implies that s_0 is Bayes against τ_0, $L(s_0 - v) = \|s_0 - v\|$ insures that τ_0 is the least favorable distribution.

Remarks. Many other well-known results for minimax decision procedures are obtainable as immediate consequences of Theorem 1. It is also well-known that when Θ is not compact, a minimax decision procedure may exist and there may be no corresponding least favorable distribution. This is a consequence of the nature of the topological dual space L^* in this case. Namely, the topological dual is a set of finitely additive set functions rather than countably additive set functions.

The next theorem provides a characterization of (v,L) optimality in terms of the extremal points of the intersection of the unit ball in L^* with K^*, where K^* is the dual of K.

As before, S is a convex set in L, a partially ordered linear space $\neq \{0\}$ of real valued functions of θ provided with a monotonic norm. Let $v \in L$ satisfy $v \leq S$ and let $d = \inf_{s \in S} \{\|s - v\|\}$. The closed unit ball in L^* will be denoted by S^* and $S^*_+ = S^* \cap K^*$. That is, S^*_+ is the set of positive linear functionals with norm not exceeding unity.

Then, we have the following theorem.

<u>Theorem 2</u>. $s_0 \in S$ is (v,L) optimal if and only if for every $s \in S$ there exists a linear functional $L_s \in L^*$ with the following properties:

(1) $\|L_s\| = 1$, (2) L_s is an extremal point of S^*_+, (3) $L_s(s) \geq L_s(s_0)$, (4) $L_s(s_0 - v) = \|s_0 - v\|$.

<u>Proof</u>. <u>Sufficiency</u>. For every $s \in S$, $\|s_0 - v\| = L_s(s_0 - v) \leq L_s(s - v) \leq \|L_s\| \, \|s - v\| = \|s - v\|$.

<u>Necessity</u>. Assume $s_0 = v$. Then since S^* is weak* compact (by Alaoglu's theorem) and K^* is weak* closed, it follows that S^*_+ is weak* compact. Also S^*_+ is convex. By the Krein-Milman theorem S^*_+ is the weak* closure of the convex hull of its extremal points. Let $E(S^*_+)$ denote these extremal points. By Theorem 1, S^*_+ contains an L with $\|L\| = 1$ and therefore $E(S^*_+)$ contains an $\tilde{L}$ with $\|\tilde{L}\| = 1$. Also $\tilde{L}(s - s_0) = \tilde{L}(s - v) \geq 0$ for all $s \in S$, since $v \leq S$. Thus for every $s \in S$, it suffices to set $L_s = \tilde{L}$.

Now assume $s_0 \neq v$. Let $A = \{L \in S^*_+ : L(s_0 - v) = \|s_0 - v\|\}$. By Theorem 1, $A \neq \phi$. Further, A is a weak* compact convex extremal subset of S^*_+. For fixed $s \in S$, let $B_s = \{L \in A : L(s - s_0) = \sup_{\tilde{L} \in A} \tilde{L}(s - s_0)\}$. Obviously, B_s is a non-empty convex weak* compact extremal subset of A. Thus there exists an extremal point L_s of B_s which necessarily is an extremal point of A and consequently an extremal point of S^*_+. Then, by Theorem 1, there is an L in A such that $L(s_0) = \inf_{s \in S} L(s)$ and therefore $0 \leq L(s - s_0) \leq \sup_{\tilde{L} \in A} \tilde{L}(s - s_0) = L_s(s - s_0)$, establishing the theorem.

The third characterization theorem uses the notion of a Kolmogorov fundamental system, which we denote by K-system.

A subset T of S^* is said to be a K-system if T is weak* closed and if for all $x \in K$, the set $T_x = \{L \in T : L(x) = \|x\|\} \neq \phi$. This leads to the following theorem.

<u>Theorem 3</u>. $s_0 \in S$ is (v,L) optimal if and only if, for any K-system T, for every $s \in S$ there is an $L_s \in T_{s_0-v}$ with $L_s(s) \geq L_s(s_0)$.

<u>Proof</u>. <u>Sufficiency</u>. For $s \in S$ and $L_s \in T_{s_0-v}$ with $L_s(s) \geq L_s(s_0)$, we have $\|s_0 - v\| = L(s_0 - v) \leq L(s - v) \leq \|s - v\|$.

<u>Necessity</u>. Now assume s_0 is (v,L) optimal and $s \in S$. Since T_{s_0-v} is weak* compact, there is an $L_s \in T_{s_0-v}$ such that $L_s(s - s_0) = \sup_{\tilde{L} \in T_{s_0-v}} L(s - s_0)$. The theorem will be established if we can show that $\sup_{\tilde{L} \in T_{s_0-v}} L(s - s_0) = \gamma_s \geq 0$. Therefore assume $\gamma_s < 0$. Let $H_s = \{L \in T : L(s - s_0) \geq 0\}$. For $L \in T_{s-v}$, $L(s_0 - v) \leq \|L\| \, \|s_0 - v\| \leq \|s_0 - v\| \leq \|s - v\| = L(s - v)$. Thus for such L, $L(s) \geq L(s_0)$ and $T_{s-v} \subset H_s$, insuring that $H_s \neq \phi$. Further, H_s is weak* compact and hence there exists an $L'_s \in H_s$ such that $L'_s(s_0 - v) = \sup_{L \in H_s} L(s_0 - v) = \alpha_s$. Now $L(s - s_0) < 0$ for all $L \in T_{s_0-v}$; consequently $T_{s_0-v} \cap H_s = \phi$. Therefore for $L \in H_s$, $L(s_0 - v) < \|s_0 - v\|$ and $\alpha_s < \|s_0 - v\|$. Also $s \neq s_0$, since otherwise we would have $L(s - s_0) = 0$ for all $L \in L^*$, but for $L \in T_{s_0-v}$, $L(s - s_0) < 0$. Therefore there exists a positive number t with $t < \min\{1, (\|s_0 - v\| - \alpha_s)/\|s - s_0\|\}$. Since $(1 - t)s_0 + ts - v \in K$ and $(1 - t)s_0 + ts \in S$, it follows that there is an $L_1 \in T$ with $L_1((1 - t)s_0 + ts - v) = \|(1 - t)s_0 + ts - v\| \geq \|s_0 - v\|$. $L_1 \notin H_s$, since $L_1(s_0 - v) + tL_1(s - s_0) \leq \alpha_s + tL_1(s - s_0) \leq \alpha_s + t\|s - s_0\| \leq \alpha_s + \|s_0 - v\| - \alpha_s$ $= \|s_0 - v\|$, a contradiction. Further, L_1 is not in $T \cap H_s^c$; since $L_1(s_0 - v) + tL_1(s - s_0) < L_1(s_0 - v) \leq \|s_0 - v\|$. Thus assuming $\gamma_s < 0$ leads to a contradiction.

Remarks. The methods used in the proofs of Theorems 1 and 2 have been substantially influenced by the work of F. R. Deutsch and P. H. Maserick [2]. The results contained therein are similar to standard theorems of approximation theory, but with one significant difference. As a consequence of the assumption $v \leq S$, the characterizing linear functionals are positive linear functionals. Theorem 3 is an adaptation of results of V. N. Nikol'skiĭ [9, 10] and was reported by G. Heindl in [6].

We conclude this section with a result, which is elementary, but nevertheless appears to be new. This gives a simple relationship between (v,L) optimality and admissibility.

Theorem 4. If $s_0 \in S$ is the unique (v,L) optimal decision procedure, then s_0 is necessarily admissible.

Proof. Assume $s_0 \in S$ is (v,L) optimal and inadmissible. Then there exists an $s_1 \in S$ with $s_1 < s_0$. Consequently, $0 \leq s_1 - v < s_0 - v$ and $\|s_1 - v\| \leq \|s_0 - v\|$. Hence, s_1 is also (v,L) optimal. Thus (v,L) optimality and inadmissibility implies non-uniqueness.

3. Concluding Remarks. The above paper constitutes an attempt to utilize some recent mathematical developments in formulating notions of statistical optimality. It is hoped that this will provide insight into statistical theory and the essential differences between various statistical philosophies.

REFERENCES

[1] D. Blackwell and M. A. Girshick, Theory of Games and Statistical Decisions, John Wiley, New York, 1954.

[2] F. R. Deutsch and P. H. Maserick, Applications of the Hahn-Banach theorem in approximation theory, SIAM Rev. 9 (1967), 516-530.

[3] N. Dunford and J. T. Schwartz, Linear Operators, Part I, Interscience Publishers, New York, 1958.

[4] T. S. Ferguson, Mathematical Statistics, A Decision Theoretic Approach, Academic Press, New York, 1967.

[5] B. Harris, Mathematical Models for Statistical Decision Theory, in Optimizing Methods in Statistics, (J. S. Rustagi, Ed.), 369-389, Academic Press, New York, 1971.

[6] G. Heindl, Ein Problem der Bestapproximation in geordneten Vektorräumen, unpublished mimeographed report.

[7] R. B. Holmes, A Course on Optimization and Best Approximation, Springer-Verlag, Berlin, 1972.

[8] V. N. Nikol'skiĭ, Extension of A. N. Kolmogorov's theorem to Banach spaces, Studies of Modern Problems of Constructive Theory of Functions (Russian), (V. I. Smirnov, Ed.), 335-337, Fizmatgiz, Moscow, 1961.

[9] V. N. Nikol'skiĭ, A characteristic property of least deviating elements in convex sets, Studies of Contemporary Problems Constructive Theory of Functions (Proc. Second All-Univ. Conf., Baku, 1962) (Russian), 80-84, Izdat. Akad. Nauk Azerbaidzan. SSR, Baku, 1965.

[10] I. Singer, Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces, Publishing House of the Academy of the Socialist Republic of Romania, Bucharest, 1970.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 1890	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE RELATION BETWEEN STATISTICAL DECISION THEORY AND APPROXIMATION THEORY		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Bernard Harris and Gerhard Heindl		8. CONTRACT OR GRANT NUMBER(s) DAAG29-75-C-0024
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS #4 - Probability, statistics and combinatorics
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE October 1978
		13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Decision theory, approximation theory.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper extends previous results by the first author (see for example, "Mathematical Models for Statistical Decision Theory" in Optimizing Methods in Statistics, J. S. Rustagi, Ed., Academic Press, 1971).

The approximation theory model describes a class of optimality principles in statistical decision theory as follows. Let S be the risk set of a statistical decision problem, that is, $S = \{R_\varphi(\theta), \theta \in \Theta, \varphi \in \Phi\}$ where Φ is the collection of randomized decision procedures, Θ is the parameter space and $R_\varphi(\theta)$ is the risk function of the statistical decision procedure φ. We interpret S as

Phi

next page

(phi

R sub phi (theta)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Theta

phi an element of Phi

Sub phi theta theta an element of Theta

phi an element of Phi

theta an element of Theta

s sub ϕ

theta

< or =

sub phi

Abstract (continued)

a set in the normed linear space L. Let $v = v(\theta)$ satisfy $v(\theta) \leq R_{\varphi}(\theta)$ for all $\varphi \in \Phi$ and all $\theta \in \Theta$. Then $s_0 \in S$ is said to be (v,L) optimal if $\|s_0 - v\| \leq \|s - v\|$ for all $s \in S$.

an element of

It is easily seen that many well-known optimality principles of statistics are of this type, such as Bayes rules and minimax rules.

In this paper, characterization theorems for this class of optimality principles are given.

abs. val. (s sub ϕ - v)

< or =

abs. val. (s - v)